

Attorney Docket # 4577-50

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of
Jung Chul LEE et al.
Serial No.: Not Yet Assigned
Filed: Concurrently herewith
For: A Text-To-Speech Conversion System
For Interlocking With Multimedia And A
Method For Organizing Input Data Of
The Same



[Handwritten signature]

LETTER TRANSMITTING PRIORITY DOCUMENT

Assistant Commissioner for Patents
Washington, D.C. 20231

SIR:

In order to complete the claim to priority in the above-identified application under 35 U.S.C. §119, enclosed herewith is a certified copy of Application No. 97-17615 filed on May 8, 1997 in Korea, upon which the priority claim is based.

Respectfully submitted,

COHEN, PONTANI, LIEBERMAN & PAVANE

By

[Handwritten signature of Martin B. Pavane]

Martin B. Pavane
Reg. No. 28,337
551 Fifth Avenue, Suite 1210
New York, N.Y. 10176
(212) 687-2770

February 9, 1998

U.S. PTO
09/020712
02/09/98

#2

대한민국 특허청
KOREAN INDUSTRIAL
PROPERTY OFFICE

별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Industrial
Property Office.

출원 번호 : 1997년 특허출원 제17615호
Application Number

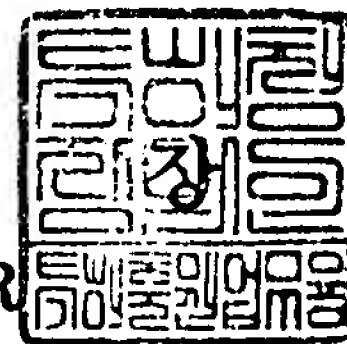
출원 년 월 일 : 1997년 5월 8일
Date of Application

출원 인 : 한국전자통신연구원
Applicant(s)

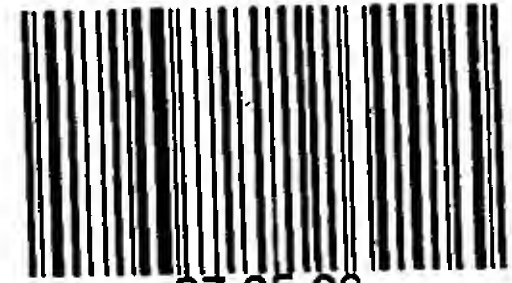
1997 년 9 월 24 일

특 허 청

COMMISSIONER



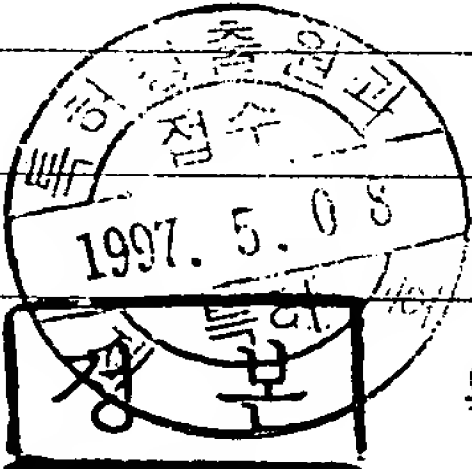
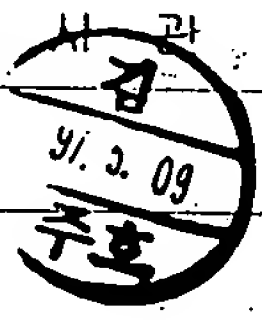
10-97-017615



97.05.08

원서번호: 1

① 3

IPC 분류 기 호	주 분 류		방 식 심 사 란	출원번호:	
	부 분 류			담	당
접 수 인 란					

특 허 출 원 서

출원인	성 명 (명칭)	한국전자통신연구원 Electronics and Telecommunications Research Institute (대표자: 양승택)				
	주민등록번호 (출원인코드)	37500300	전화 번호	042-860-6506	국 적	대한민국
	주 소	대전광역시 유성구 가정동 161 (305-350)				
대리인	성 명	최승민	대리인 코드	L112	전화 번호	02-3149-7805
	주 소	서울특별시 중구 순화동 1-170 삼도빌딩 4층 (100-130)				
	성 명	신영무	대리인 코드	G041	전화 번호	02-3149-7805
	주 소	서울특별시 중구 순화동 1-170 삼도빌딩 4층 (100-130)				
발명자	성 명	이정철 LEE JUNG CHUL,				
	주민등록번호	611007-1108615	국 적	대한민국		
	주 소	대전광역시 중구 태평동 삼부아파트 36-102 (301-150)				
	성 명	한민수 HAHN MIN SOO,				
	주민등록번호	561123-1047916	국 적	대한민국		
	주 소	대전광역시 유성구 신성동 삼성한울아파트 106-1004 (305-345)				
	성 명	이항섭 LEE HANG SEOP,				
	주민등록번호	670312-1011215	국 적	대한민국		
주 소	대전광역시 서구 만년동 초원아파트 106-1509 (302-150)					

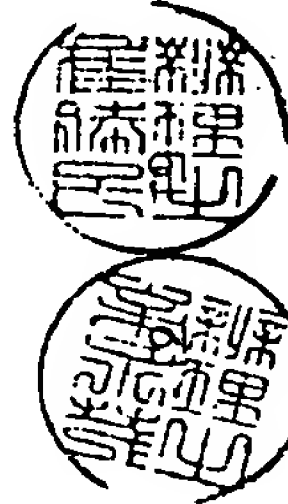
발명의 명칭	다중매체와의 연동을 위한 텍스트/음성변환기 및 그 입력 데이터 구조화 방법
	Syntax for TTS input data to synchronize with multimedia

특허법 제42조의 규정에 의하여 위와 같이 출원합니다.

1997년 05월 08일

대리인 최승민

신영무



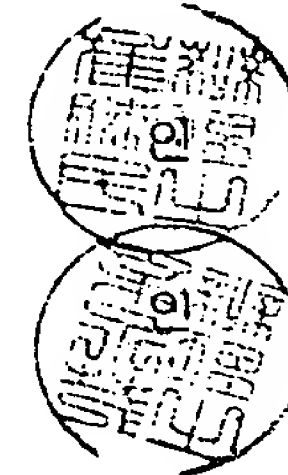
특허청장 귀하

특허법 제60조의 규정에 의하여 위와 같이 출원심사를 청구합니다.

1997년 05월 08일

대리인 최승민

신영무



특허청장 귀하

* 첨부서류	수 수 료			
	출원료	기본	20 면	22000 원
1. 요약서, 명세서(및 도면) 각 1통 2. 출원서 부분, 요약서, 명세서(및 도면)을 포함하는 FD부분 1통 3. 위임장(및 동 번역문)		가산	3 면	2400 원
	우선권 주장료		0 건	0 원
	심사 청구료		12 항	292000 원
	합 계			316400 원

【요약서】

【요약】

본 발명은 다중매체 환경에서 텍스트/음성변환기(text-to-speech conversion system; TTS) 연동방법에 있어서 텍스트/음성변환기용 입력데이터 구조화 방법에 관한 것이다.

기존의 합성기는 입력된 텍스트로부터 음성을 합성하는 용도로만 고려되고 있는 상황이다. 그런데 텍스트/음성변환기를 이용하여 동영상에 더빙을 하고자 할 때나, 애니메이션과 같은 다중매체와 합성음 간의 자연스러운 연동을 구현하기 위해서 필요한 동기화 정보는 단지 텍스트로부터 추정하기가 불가능하며, 합성음의 자연성 향상을 위한 부가 데이터 사용, 그리고 이들 데이터의 구조화에 대한 연구결과는 거의 없는 실정이다.

따라서, 본 발명은 텍스트/음성변환기에서 텍스트 이외에 부가적 운율 정보, 다중매체와의 연동에 필요한 정보, 그리고 이들 정보와 텍스트/음성변환기 간의 인터페이스를 정의하여 합성음 생성에 사용함으로써 합성음의 자연성 향상과 다중매체와 TTS간의 동기화 구현을 그 목적으로 한다.

본 발명은 실제 음성데이터와 동영상의 입술모양을 분석하여 추정된 운율정보, 입술모양 정보와 텍스트 정보의 구조화 및 합성기 응용을 통해 합성음의 자연성과 동영상과의 동기화를 구현함으로써 외화등에 한국어 더빙, 통신 서비스, 사무 자동화, 교육 등의 여러 분야에 응용할 수 있는 탁월한 효과가

있다.

【대표도】

도 3.

【명세서】

【발명의명칭】

다중매체와의 연동을 위한 텍스트/음성변환기 및 그 입력 데이터 구조화 방법

【도면의간단한설명】

도 1은 종래의 텍스트/음성변환기의 구성도.

도 2는 본 발명이 적용되는 하드웨어의 구성도.

도 3은 본 발명에 따른 텍스트/음성변환기의 구성도.

<도면의 주요 부분에 대한 부호의 설명>

- | | |
|-----------------|-------------------|
| 1: 언어 처리부 | 2: 운율 처리부 |
| 3: 신호 처리부 | 4: 합성 단위 데이터 베이스 |
| 5: 데이터 입력 장치 | 6: 중앙 처리 장치 |
| 7: 합성 데이터 베이스 | 8: D/A 변환 장치 |
| 9: 영상 출력 장치 | 10: 다중 매체 입력 정보 |
| 11: 매체별 데이터 분배기 | 12: 언어 처리부 |
| 13: 운율 처리부 | 14: 동기 조정기 |
| 15: 신호 처리부 | 16: 합성 단위 데이터 베이스 |
| 17: 영상 출력 장치 | |

【발명의상세한설명】

【발명의목적】

【발명이속하는기술분야및그분야의종래기술】

본 발명은 다중매체와 연동을 위한 텍스트/음성변환기(text-to-speech conversion system; 이하 TTS라 칭함) 및 그 입력 데이터 구조화 방법에 관한 것으로, 특히 TTS 에서 텍스트 이외에 부가적 운율정보, 다중매체와의 연동에 필요한 정보, 그리고 이들 정보와 텍스트/음성변환기 간의 인터페이스를 정의하여 합성음 생성에 사용함으로써, 합성음의 자연성 향상과 다중매체와 TTS간의 동기화 할 수 있는 다중매체와의 연동을 위한 텍스트/음성변환기 및 그 입력 데이터 구조화 방법에 관한 것이다.

일반적으로, 음성합성기의 기능은 컴퓨터가 사용자인 인간에게 다양한 형태의 정보를 음성으로 제공하는데 있다. 이를 위해서 음성합성기는 사용자에게 주어진 텍스트로부터 고품질의 음성 합성 서비스를 제공할 수 있어야 한다. 뿐만 아니라 동영상이나 애니메이션 등의 다중매체 환경에서 제작된 데이터 베이스나 대화 상대방으로부터 제공되는 다양한 미디어와 연동되기 위해서는 이들 미디어와 동기화 되도록 합성음을 생성할 수 있어야 한다. 특히 다중매체와 TTS간의 동기화는 사용자에게 고품질의 서비스를 제공하기 위해 필수적이다.

기존의 TTS는 도 1에 도시된 바와 같이 입력된 텍스트로부터 합성음을 생성하기까지 일반적으로 3 단계의 과정을 거치게 된다.

1 단계인 언어 처리부(1)에서는 텍스트를 음소열로 변환하고, 운율 정

보를 추정하여 이를 심볼화 한다. 운율 정보의 심볼은 구문구조 분석결과를 이용한 구.절 경계, 단어내 엑센트 위치, 문형 등으로부터 추정된다.

2 단계인 운율 처리부(2)는 심볼화된 운율 정보로부터 규칙 및 테이블을 이용하여 운율 제어 파라미터의 값을 계산한다. 운율 제어 파라미터로는 음소의 지속시간, 피치 형태(contour), 에너지 형태(contour), 쉼 구간 정보가 있다.

3 단계인 신호처리부(3)는 합성 단위 데이터 베이스(4)와 운율 제어 파라미터를 이용하여 합성음을 생성한다. 즉 기존의 합성기는 언어 처리부(1)와 운율 처리부(2)에서 자연성, 발성 속도와 관련된 정보를 단지 입력 텍스트만으로 추정을 해야 함을 의미한다.

또한, 기존의 TTS는 문장 단위로 입력된 데이터를 합성음으로 출력하는 단순한 기능을 가지고 있다. 그러므로 파일내 저장된 문장, 혹은 통신망을 통해 입력된 문장들을 연속해서 합성음으로 출력하기 위해서는 입력 데이터에서 문장을 읽어서 TTS의 입력으로 전달하는 주 제어 프로그램이 필요하다. 이러한 주 제어 프로그램 중에는 입력된 데이터로부터 텍스트를 분리하여 단순히 처음부터 끝까지 1회 합성음을 출력하는 방법, 텍스트 편집기와 연동하여 합성음을 생성하는 방법, 그래픽 인터페이스를 이용하여 문장을 검색하고 합성음을 생성하는 방법 등이 있지만 그 대상은 텍스트로 제한되어 있다.

현재 TTS에 대한 연구가 세계 여러 나라에서 자국어에 대상으로 많이 진행되어 일부 상용화가 이루어졌다. 그러나, 아직 입력된 텍스트로부터 음성

을 합성하는 용도로만 고려되고 있는 상황이다. 그런데 TTS를 이용하여 동영상에 더빙을 하고자 할 때나, 애니메이션과 같은 다중매체와 합성음 간의 자연스러운 연동을 구현하기 위해서 필요한 동기화 정보는 단지 텍스트로부터 추정하기는 불가능하므로 현재의 구조로는 이들 기능을 구현할 수 있는 방법이 없다. 또한, 합성음의 자연성 향상을 위한 부가 데이터 사용, 그리고 이들 데이터의 구조화에 대한 연구결과는 없는 실정이다.

【발명이 이루고자 하는 기술적 과제】

따라서, 본 발명은 TTS 에서 텍스트 이외에 부가적 운율정보, 다중매체와의 연동에 필요한 정보, 그리고 이들 정보와 TTS 간의 인터페이스를 정의하여 합성음 생성에 사용함으로써, 합성음의 자연성 향상과 다중매체와 TTS 간의 동기화 할 수 있는 다중매체와의 연동을 위한 텍스트/음성변환기의 구조 및 입력 데이터 구조화 방법을 제공하는 데 그 목적이 있다.

상술한 목적을 달성하기 위한 본 발명에 따른 다중매체와의 연동을 위한 텍스트/음성변환기는 텍스트, 운율, 동화상과의 동기화 정보, 입술모양 및 개인성 등의 정보를 구조화 시키는 다중 매체 정보 입력부와, 상기 다중 매체 정보 입력부의 정보를 매체별 정보로 분리하는 매체별 데이터 분배기와, 상기 매체별 데이터 분배기로 부터 분배된 텍스트를 음소열로 변환하고, 운율 정보를 추정하여 이를 심볼화하는 언어 처리부와, 상기 심볼화된 운율 정보로부터 규칙 및 테이블을 이용하여 운율 제어 파라미터의 값을 계산하는 운율 처리부

와, 상기 매체별 데이터 분배기로 부터 분배된 동기화 정보를 이용하여 음소의 지속시간을 조정하는 동기 조정기와, 상기 운율 제어 파라미터와 합성 단위 데이터 베이스 내의 데이터를 이용하여 합성음을 생성하는 신호처리부와, 상기 매체별 데이터 분배기로 부터 분배된 영상 정보를 화면에 출력하는 영상 출력 장치로 구성된 것을 특징으로 한다.

상술한 목적을 달성하기 위한 본 발명에 따른 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법은 다중 매체 정보 입력부에서 합성음의 자연성 향상과 다중매체와 텍스트/음성변환기 간의 동기화 구현을 위해 구조화된 다중 매체 입력 정보의 구성을 텍스트, 운율, 동화상과의 동기화 정보, 입술모양, 개인성 정보로 구분하는 단계와, 상기 다중 매체 정보 입력부에서 구분된 정보를 매체별 데이터 분배기에서 각각의 정보에 따라 분배하도록 하는 단계와, 상기 매체별 데이터 분배기에서 분배된 텍스트를 언어 처리부에서 음소열로 변환하고, 운율 정보를 추정하여 이를 심볼화 하는 단계와, 상기 운율 정보를 운율 처리부에서 다중 매체 정보에 포함되어 있는 운율제어 파라미터 이외의 운율 제어 파라미터의 값을 계산하는 단계와, 상기 운율 처리부의 처리 결과를 동기 조정기에서 동기화 정보의 입력에 따라 영상신호와의 동기를 맞추기 위해 음소별 지속시간을 조정하는 단계와, 상기 매체별 데이터 분배기로부터의 운율 정보 및 상기 동기 조정기의 처리결과를 신호처리부에서 합성 단위 데이터 베이스를 이용하여 합성음을 생성하는 단계와, 상기 매체별 데이터 분배기로부터 분배된 영상 정보를 영상 출력 장치에서 화면에 출력 하

도록 하는 단계로 이루어진 것을 특징으로 한다.

【발명의구성및작용】

이하, 첨부된 도면을 참조하여 본 발명을 상세히 설명하기로 한다.

도 2는 본 발명이 적용되는 하드웨어의 구성도이다. 다중 데이터 입력 장치(5), 중앙 처리 장치(6), 합성 데이터 베이스(7), 디지털/아날로그(D/A) 변환 장치(8) 및 영상 출력 장치(9)로 구성된다.

다중 데이터 입력 장치(5)는 영상, 텍스트 등의 다중 매체로 구성된 데이터를 입력받아 중앙 처리 장치(6)로 출력한다.

중앙 처리 장치(6)는 본 발명의 다중 데이터 입력을 분배하고 동기를 조정하며 합성음을 생성하는 알고리즘이 탑재되어 수행된다.

합성 데이터베이스(7)는 합성 알고리즘에 사용되는 합성 데이터 베이스로서 기억장치에 저장되어 있으며 상기 중앙 처리 장치(6)로 필요한 데이터를 전송한다.

디지털/아날로그(D/A) 변환장치(8)는 합성이 끝난 디지털 데이터를 아날로그 신호로 변환하여 외부로 출력한다.

영상 출력 장치(9)는 입력된 영상정보를 화면에 출력한다.

표 1 및 표 2는 본 발명에 적용되는 구조화된 다중 매체 입력 정보 상태를 나타내는 알고리즘으로서, 텍스트, 운율, 동화상과의 동기화 정보, 입술 모양, 개인성 정보로 이루어져 있다.

【표 1】

Syntax
<pre> TTS_Sequence() { TTS_Sequence_Start_Code TTS_Sentence_ID Language_Code Prosody_Enable Video_Enable Lip_Shape_Enable Trick_Mode_Enable do{ TTS_Sentence() }while(next_bits()==TTS_Sentence_Start_Code } </pre>

여기서, TTS_Sequence_Start_Code는 Hexadecimal 'XXXXXX'로 표시된 bit string으로서 TTS 데이터 열의 시작을 의미한다.

TTS_Sentence_ID는 10-bit ID로서 각 TTS 데이터 열의 고유번호를 나타낸다.

Language_Code는 한국어, 영어, 독일어, 일어, 프랑스어 등과 같이 합성하고자 하는 대상 언어를 나타낸다.

Prosody_Enable은 1-bit flag로서 원음의 운율 데이터가 구조화 데이터에 포함되면 1의 값을 갖는다.

Video_Enable은 1-bit flag로서 TTS가 동영상과 연동될 때 1의 값을 가진다.

Lip_Shape_Enable은 1-bit flag로서 입술모양 데이터가 구조화 데이터에 포함되면 1의 값을 가진다.

Trick_Mode_Enable은 1-bit flag로서 stop, restart, forward, backward와 같은 trick mode를 지원하도록 데이터가 구조화 되면 1의 값을 가진다.

Syntax
<pre> TTS_Sentence() { TTS_Sentence_Start_Code TTS_Sentence_ID Silence if(Silence) { Silence_Duration } else { Gender Age if(!Video_Enable) { Speech_Rate } Length_of_Text TTS_Text() if(Prosody_Enable) { Dur_Enable F0_Contour_Enable Energy_Contour_Enable Number_of_Phonemes for(j=0 ; j<Number_of_phonemes ; j++) { Symbol_each_phoneme if(Dur_Enable) { Dur_each_phoneme } if(F0_Contour_Enable) { F0_contour_each_phoneme } if(Energy_Contour_Enable) { Energy_contour_each_phoneme } } } if(Video_Enable) { Sentence_Duration Position_in_Sentence offset } if(Lip_Shape_Enable) { Number_of_Lip_Event for(j=0 ; j<Number_of_Lip_Event ; j++) { Lip_in_Sentence Lip_shape } } } } } </pre>

여기서, TTS_Sentence_Start_Code는 Hexadecimal 'XXXXXX'로 표시된 bit string으로서 TTS 문장의 시작을 의미한다. 10-bit ID로서 각 TTS 데이터 열의 고유번호를 나타낸다.

TTS_Sentence_ID는 10-bit ID로서 TTS 열내 각 TTS 문장의 고유번호를 나타낸다.

Silence는 1-bit flag 현재 입력 프레임이 무음구간일 때 '1'이 된다.

Silence_Duration은 현 무음구간의 지속 시간을 milliseconds로 나타낸다.

Gender는 1-bit로 합성음의 남녀 성별을 구분한다.

Age는 합성음의 나이를 유아, 청소년, 중년, 노년으로 구분한다.

Speech_Rate는 합성음의 발성 속도를 나타낸다.

Length_of_Text는 입력 텍스트의 문장의 길이를 byte로 나타낸다.

TTS_Text는 임의의 길이 문장 텍스트를 나타낸다.

Dur_Enable은 1-bit flag로서 각 음소의 지속시간 정보가 구조화 데이터에 포함될 때 '1'이 된다.

F0_Contour_Enable은 1-bit flag로서 각 음소의 피치 정보가 구조화 데이터에 포함될 때 '1'이 된다.

Energy_Contour_Enable은 1-bit flag로서 각 음소의 에너지 정보가 구조화 데이터에 포함될 때 '1'이 된다.

Number_of_Phonemes는 문장의 합성에 필요한 음소의 수를 나타낸다.

Symbol_each_phoneme은 IPA와 같은 각 음소를 나타내는 심볼을 나타낸다.

Dur_each_phoneme은 음소의 지속시간을 표시한다.

F0_contour_each_phoneme은 음소의 피치 패턴을 나타내는 것으로 음소의 시작점, 중간, 끝점에서의 피치값으로 표시한다.

Energy_contour_each_phoneme은 음소의 에너지 패턴을 나타내는 것으

로 음소의 시작점, 중간, 끝점에서의 에너지 값을 dB로 표시한다.

Sentence_Duration은 문장에 대한 합성음의 전체 지속시간을 나타낸다.

Position_in_Sentence는 현재 프레임의 문장내 위치를 나타낸다.

offset은 동영상과 연동되는 경우, GOP(Group Of Pictures) 내에 문장의 시작점이 있을 때 GOP 시작점으로부터 문장의 시작점까지의 지연시간을 나타낸다.

Number_of_Lip_Event는 문장내 입술모양 변화점들의 개수를 나타낸다.

Lip_in_Sentence는 문장내 입술모양 변화점의 위치를 나타낸다.

Lip_shape는 문장내 입술모양 변화점에서 입술모양을 나타낸다.

텍스트 정보는 사용언어에 대한 분류코드, 문장 텍스트를 포함한다. 운율정보에는 문장내 음소의 수, 음소열 정보, 음소별 지속시간, 음소의 피치 패턴, 음소의 에너지 패턴이 있으며 합성음의 자연성을 향상시키는데 사용한다. 동화상과 합성음의 동기화 정보는 더빙의 개념으로 살펴볼 때, 그 구현 방식에 따라 3가지 경우로 나눌 수 있다.

첫째로는 문장단위로 동화상과 합성음을 동기화 시키는 방법으로서 문장의 시작점, 지속시간, 시작점 지연시간 정보를 이용하여 합성음의 지속시간을 조절한다. 각 문장의 시작점은 동영상내에서 각 문장에 대한 합성음의 출력이 시작되어야 할 장면들의 위치를 나타내며, 문장의 지속시간은 각 문장에 대한 합성음이 지속되는 장면 수를 표시한다. 그리고, 그룹영상(Group of Picture: GOP) 개념이 이용되는 MPEG-2, MPEG-4 영상압축 방식의 동화상은 재

생시 임의의 장면에서부터 시작할 수 없고 반드시 그룹영상내 시작 장면에서부터 재생하게 되어 있다. 그러므로 시작점 지연시간은 그룹영상과 TTS가 동기를 맞추기 위해 필요한 정보이고, 그룹영상내 시작 장면과 발성 시작점 간의 지연시간을 나타낸다. 이 방법은 구현이 쉽고 부가적 노력이 최소화되는 장점이 있지만, 자연스러운 동기화와는 거리가 멀다.

두번째 방법으로는 동영상에서 음성신호와 관련된 구간에서는 매 음소마다 시작점, 끝점 정보와 음소 정보를 표기하여 이 정보를 합성음 생성에 이용하는 방법이다. 이 방법은 음소단위로 동화상과 합성음의 동기를 맞출 수 있으므로 정확도가 높은 장점이 있지만 동화상의 음성구간에서 음소단위로 지속시간 정보를 검출하여 기록하기 위한 부가적 노력이 아주 많은 단점이 있다.

세번째 방법으로는 음성의 시작점, 끝점, 입술의 모양, 입술모양의 변화시점을 기준으로 하여 동기화 정보를 기록하는 방법이다. 입술모양은 입술상하간의 거리 (열림 정도), 입술 좌우 끝점간의 거리 (벌림 정도), 입술의 내뺄 정도로 수치화하며, 변별적 특성이 높은 패턴을 기준으로 음소의 조음위치, 조음방법에 따라 입술 모양을 정량화, 정규화된 패턴으로 정의한다. 이 방법은 동기화를 위한 정보 제작의 부가적 노력을 최소화하면서 동기화 효율을 높이는 방법이다.

본 발명에 적용되는 구조화된 다중 매체 입력 정보는 이상의 3가지 동기화 방식을 정보제공자가 임의로 선택하여 구현할 수 있게 해준다.

또한, 입술 애니메이션을 구현하는 방법에도 구조화된 입력정보를 이용

한다. 입력된 텍스트로부터 TTS에서 작성한 음소열과 음소별 지속시간, 혹은 입력정보에서 분배된 음소열과 음소별 지속시간을 이용하여 입술 애니메이션을 구현할 수도 있고, 입력정보에 포함된 입술모양 정보를 이용하여 입술 애니메이션을 구현할 수도 있다.

개인성 정보는 사용자가 합성음의 성별, 연령, 합성음 발성속도의 변화 등을 가능하게 한다. 성별에는 남, 여가 있고 연령별에는 6-7세, 18세, 40세, 65세 정도의 4가지로 분류한다. 발성속도의 변화는 표준속도의 0.7배에서 1.6배의 10단계로 변화를 줄 수 있다. 이들 정보를 이용하여 합성음의 음질을 다양화한다.

도 3은 본 발명에 따른 텍스트/음성변환기의 구성도이다. 다중 매체 정보 입력부(10), 매체별 데이터 분배기(11), 표준화된 언어 처리부(12), 운율 처리부(13), 동기 조정기(14), 신호처리부(15), 합성 단위 데이터 베이스(16) 및 영상 출력 장치(17)로 구성된다.

다중 매체 정보 입력부(10)는 표 1 및 표 2의 형식으로 구성되어 있는데 텍스트, 운율정보, 동화상과의 동기화 정보, 입술모양 정보로 이루어져 있다. 이중 필수 정보는 텍스트이고, 기타 정보는 개인성과 자연성 향상과 다중 매체와의 동기화를 위한 선택 사항으로서 정보제공자가 선택적으로 제공할 수 있으며, 필요시 TTS 사용자가 문자입력장치나 마우스를 이용하여 수정이 가능하다. 이들 정보는 다중 매체 분배기(11)에 전달된다.

다중 매체 분배기(11)는 다중 매체 정보를 전달받아서 영상 정보는 영

상 출력 장치(17)로 전달하고, 텍스트는 언어 처리부(12)로 전달하며, 동기화 정보는 동기 조정기(14)에서 사용할 수 있는 데이터 구조로 변환하여 전달한다. 입력된 다중 매체 정보내에 운율정보가 있으면 신호처리부에서 사용할 수 있는 데이터 구조로 변환하여 운율 처리부와 동기 조정기(14)로 전달하며, 개인성 정보가 있으면 TTS 내에 합성단위 데이터 베이스와 운율 처리부에서 사용할 수 있는 데이터 구조로 변환하여 전달한다.

언어 처리부(12)는 텍스트를 음소열로 변환하고, 운율 정보를 추정하여 이를 심볼화 한 뒤 운율 처리부(13)에 보낸다. 운율 정보의 심볼은 구문구조 분석결과를 이용한 구.절 경계, 단어내 엑센트 위치, 문형 등으로부터 추정된다.

운율 처리부(13)는 상기 언어 처리부(12)의 처리 결과를 받아서 다중 매체 정보에 포함되어 있는 운율제어 파라미터 이외의 운율 제어 파라미터의 값을 계산한다. 운율 제어 파라미터로는 음소의 지속시간, 피치 contour, 에너지 contour, 쉼 위치 및 길이가 있다. 계산된 결과는 동기 조정기(15)로 전달된다.

동기 조정기(14)는 상기 운율 처리부(13)의 처리 결과를 받아서 영상 신호와의 동기를 맞추기 위해 음소별 지속시간을 조정한다. 음소별 지속시간의 조정은 매체별 데이터 분배기(11)에서 보내온 동기화 정보를 이용한다. 먼저 각 음소별 조음장소, 조음방법에 따라 입술모양을 각 음소에 할당하고 이를 토대로 동기화 정보에 있는 입술모양과 비교하여 음소열을 동기화 정보에 기록된

입술모양 갯수만큼 소 그룹으로 분리한다. 그리고 소 그룹내의 음소 지속시간은 동기화 정보에 포함되어 있는 입술모양의 지속시간 정보를 이용하여 다시 계산한다. 조정된 지속시간 정보를 운율 처리부의 결과에 포함시켜 신호처리부(15)로 전달한다.

신호처리부(15)는 다중 매체 분배기(11)로부터 운율정보를 받거나 상기 동기 조정기(14)의 처리결과를 받아서 합성 단위 데이터 베이스(16)를 이용하여 합성음을 생성하여 출력한다.

합성 단위 데이터 베이스(16)는 다중 매체 분배기(11)로부터 개인성 정보를 받아서 성, 연령에 적합한 합성 단위들을 선정한 뒤 신호처리부(15)의 요구를 받아서 합성에 필요한 데이터를 신호처리부(15)로 전송한다.

【발명의효과】

상술한 바와 같이 본 발명은 실제 음성데이터를 분석하여 추정된 개인성, 운율 정보를 텍스트 정보와 함께 다단계 정보로 구성하고, 합성음 생성에 직접 이용함으로써, 합성음의 개인성을 구현하고 자연성을 향상시키며, 실제 음성데이터와 동영상의 입술모양을 분석하여 추정된 입술모양 정보와 텍스트 정보를 합성음 생성에 직접 이용하는 방식을 통해 합성음과 동영상과의 동기화를 구현함으로써, 외화등에 한국어 더빙을 가능하게 하고, 다중 매체 환경에서 영상정보와 TTS의 동기화를 가능하게 함으로써, 통신 서비스, 사무 자동화, 교육 등의 여러 분야에 응용할 수 있는 탁월한 효과가 있다.

【특허청구의범위】

【청구항 1】

텍스트, 운율, 동화상과의 동기화 정보, 입술모양 및 개인성 등의 정보를 구조화 시키는 다중 매체 정보 입력부와,

상기 다중 매체 정보 입력부의 정보를 매체별 정보로 분리하는 매체별 데이터 분배기와,

상기 매체별 데이터 분배기로 부터 분배된 텍스트를 음소열로 변환하고, 운율 정보를 추정하여 이를 심볼화하는 언어 처리부와,

상기 심볼화된 운율 정보로부터 규칙 및 테이블을 이용하여 운율 제어 파라미터의 값을 계산하는 운율 처리부와,

상기 매체별 데이터 분배기로 부터 분배된 동기화 정보를 이용하여 음소의 지속시간을 조정하는 동기 조정기와,

상기 운율 제어 파라미터와 합성 단위 데이터 베이스 내의 데이터를 이용하여 합성음을 생성하는 신호처리부와,

상기 매체별 데이터 분배기로 부터 분배된 영상 정보를 화면에 출력하는 영상 출력 장치로 구성된 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기.

【청구항 2】

다중 매체 정보 입력부에서 합성음의 자연성 향상과 다중매체와 텍스트

/음성변환기 간의 동기화 구현을 위해 구조화된 다중 매체 입력 정보로 구분하는 단계와,

상기 다중 매체 정보 입력부에서 구분된 정보를 매체별 데이터 분배기에서 각각의 정보에 따라 분배하도록 하는 단계와,

상기 매체별 데이터 분배기에서 분배된 텍스트를 언어 처리부에서 음소열로 변환하고, 운율 정보를 추정하여 이를 심볼화 하는 단계와,

상기 운율 정보를 운율 처리부에서 다중 매체 정보에 포함되어 있는 운율 제어 파라미터 이외의 운율 제어 파라미터의 값을 계산하는 단계와,

상기 운율 처리부의 처리 결과를 동기 조정기에서 동기화 정보의 입력에 따라 영상신호와의 동기를 맞추기 위해 음소별 지속시간을 조정하는 단계와,

상기 매체별 데이터 분배기로부터의 운율 정보 및 상기 동기 조정기의 처리결과를 신호처리부에서 합성 단위 데이터 베이스를 이용하여 합성음을 생성하는 단계와,

상기 매체별 데이터 분배기로부터 분배된 영상 정보를 영상 출력 장치에서 화면에 출력하도록 하는 단계로 이루어진 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【청구항 3】

제 2 항에 있어서, 상기 구조화된 다중 매체 정보는 텍스트 정보, 운율

정보, 동영상과의 동기화 정보, 입술모양 및 개인성 정보로 이루어진 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【청구항 4】

제 3 항에 있어서, 상기 운율 정보는 문장내 음소의 수, 음소열 정보, 음소별 지속시간, 음소의 피치 패턴 및 음소의 에너지 패턴으로 이루어진 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【청구항 5】

제 4 항에 있어서, 상기 음소의 피치 패턴은 음소내 시작점, 중간점, 끝점에서의 피치값으로 표시하는 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【청구항 6】

제 4 항에 있어서, 상기 음소의 에너지 패턴은 음소내 시작점, 중간점, 끝점에서의 에너지값을 데시벨 값으로 표시하는 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【청구항 7】

제 2 항에 있어서, 상기 동기화 정보는 텍스트, 입술모양, 동화상내 위치 정보, 지속시간 정보로 이루어진 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【청구항 8】

제 2 항에 있어서, 상기 동기화 정보를 문장의 시작점 지속시간 및 시작점 지연시간 정보로 구성하고 이 정보를 이용하여 음소별 지속시간을 조정하는 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【청구항 9】

제 2 항에 있어서, 상기 동기화 정보를 문장의 시작점 지속시간 및 시작점 지연시간 정보로 구성하고, 이 정보를 이용하여 텍스트내 음소의 조음 방법, 조음조절을 고려한 예측 입술모양과 동기 정보내의 입술모양 및 지속시간 정보로 구성하여 음소별 지속시간을 조정하는 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【청구항 10】

제 2 항에 있어서, 상기 합성음 생성은 동영상에서 음성신호와 관련된

구간에서 매 음소마다 시작점, 끝점 정보와 음소 정보를 이용해 생성하는 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【청구항 11】

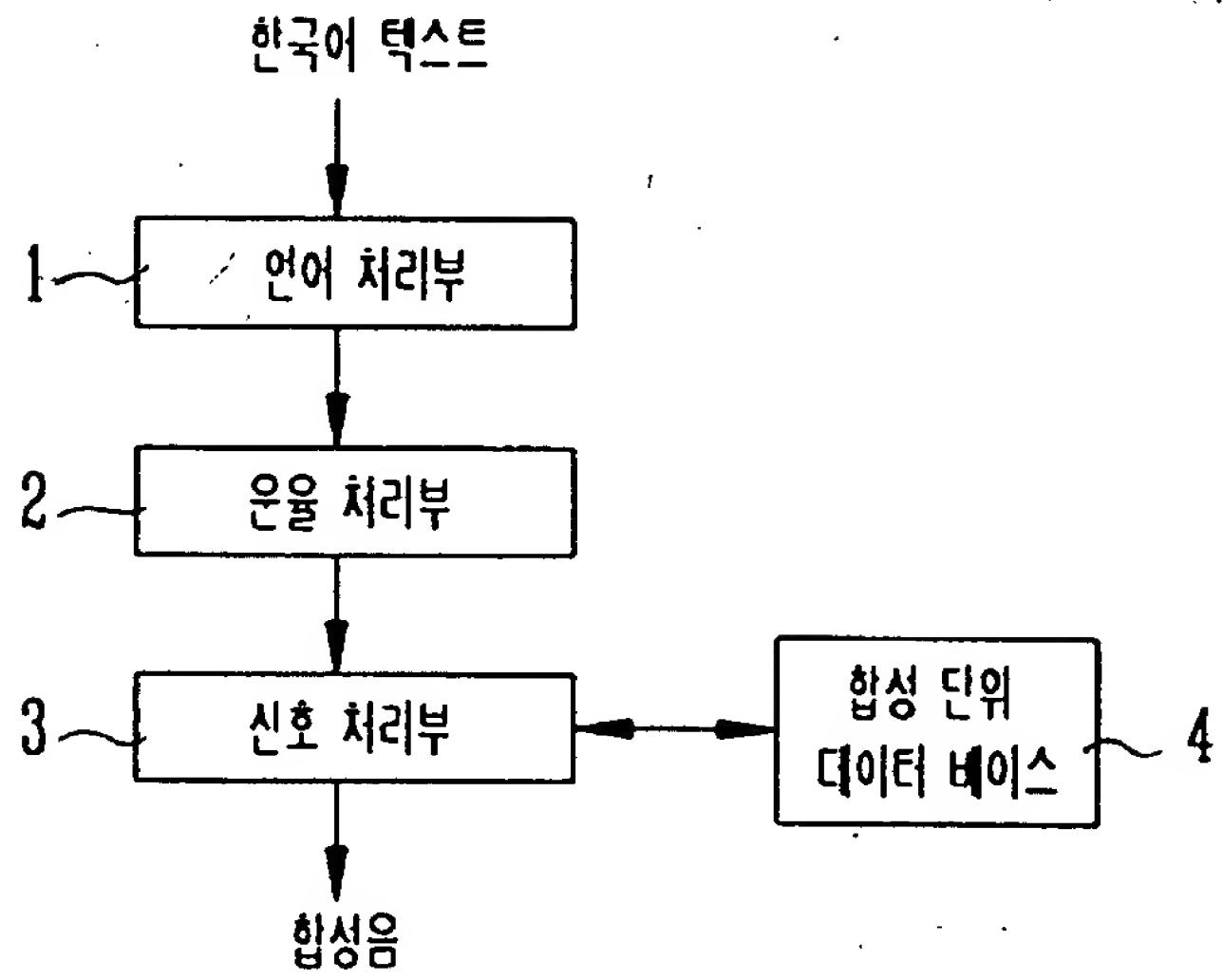
제 2 항에 있어서, 상기 합성음 생성은 입술모양을 입술 모양의 상하간 거리(열림정도), 입술 좌우 끝점간의 거리(벌림정도), 입술의 내뺨정도를 수치화 하며, 변별적 특성이 높은 패턴을 기준으로 음소의 조음 위치, 조음 방법에 따라 입술모양을 정량화, 정규화된 패턴으로 생성하는 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【청구항 12】

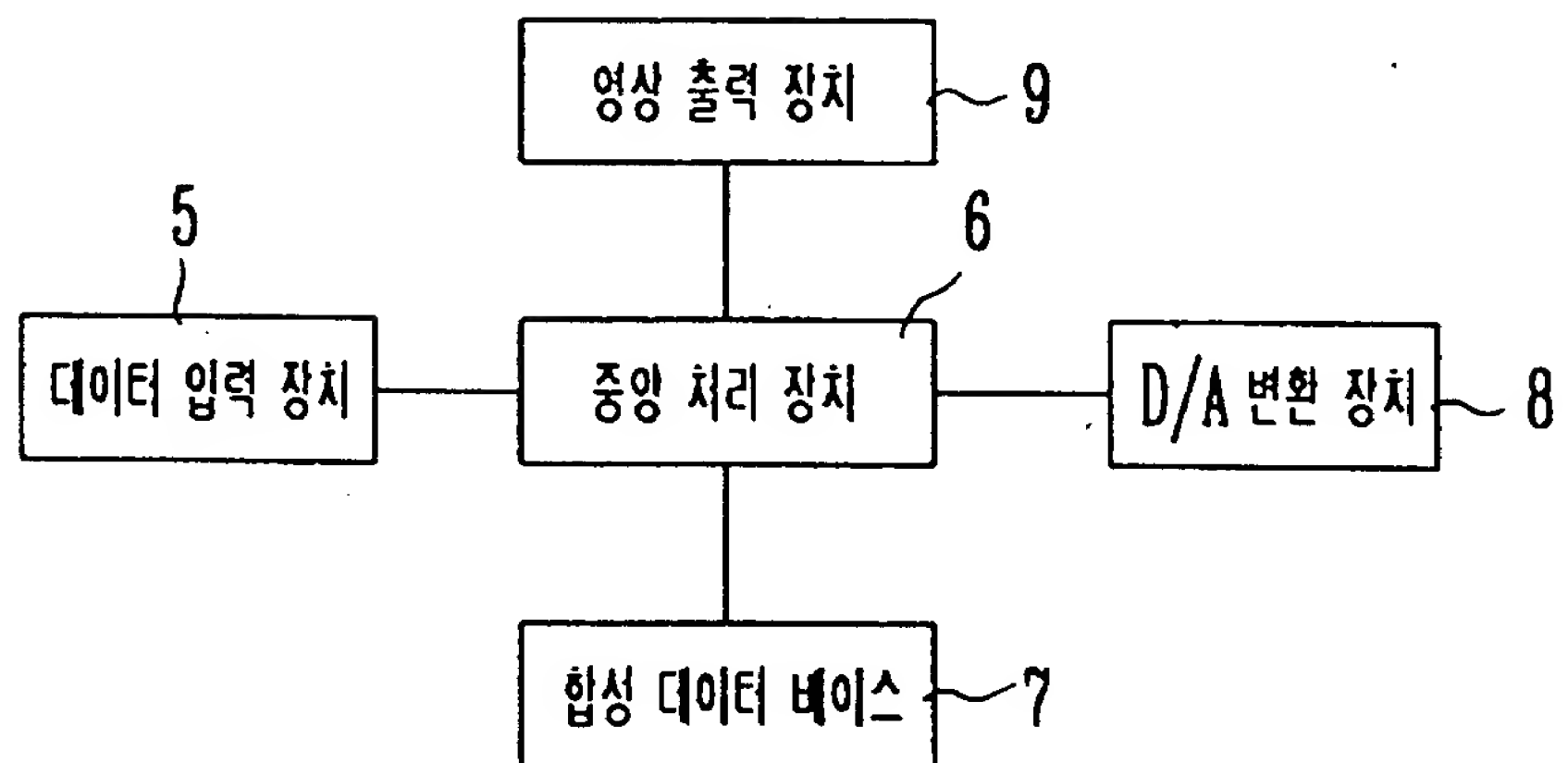
제 2 항에 있어서, 상기 매체 정보 전달 방법은 매체 정보내에 운율정보가 있으면 신호처리부에서 사용할 수 있는 데이터 구조로 변환하여 운율 처리부와 동기 조정기로 전달하고, 개인성 정보가 있으면 텍스트/음성변환기 내에 합성단위 데이터 베이스와 운율 처리부에서 사용할 수 있는 데이터 구조로 변환하여 전달 하도록 하는 것을 특징으로 하는 다중매체와의 연동을 위한 텍스트/음성변환기의 입력 데이터 구조화 방법.

【도면】

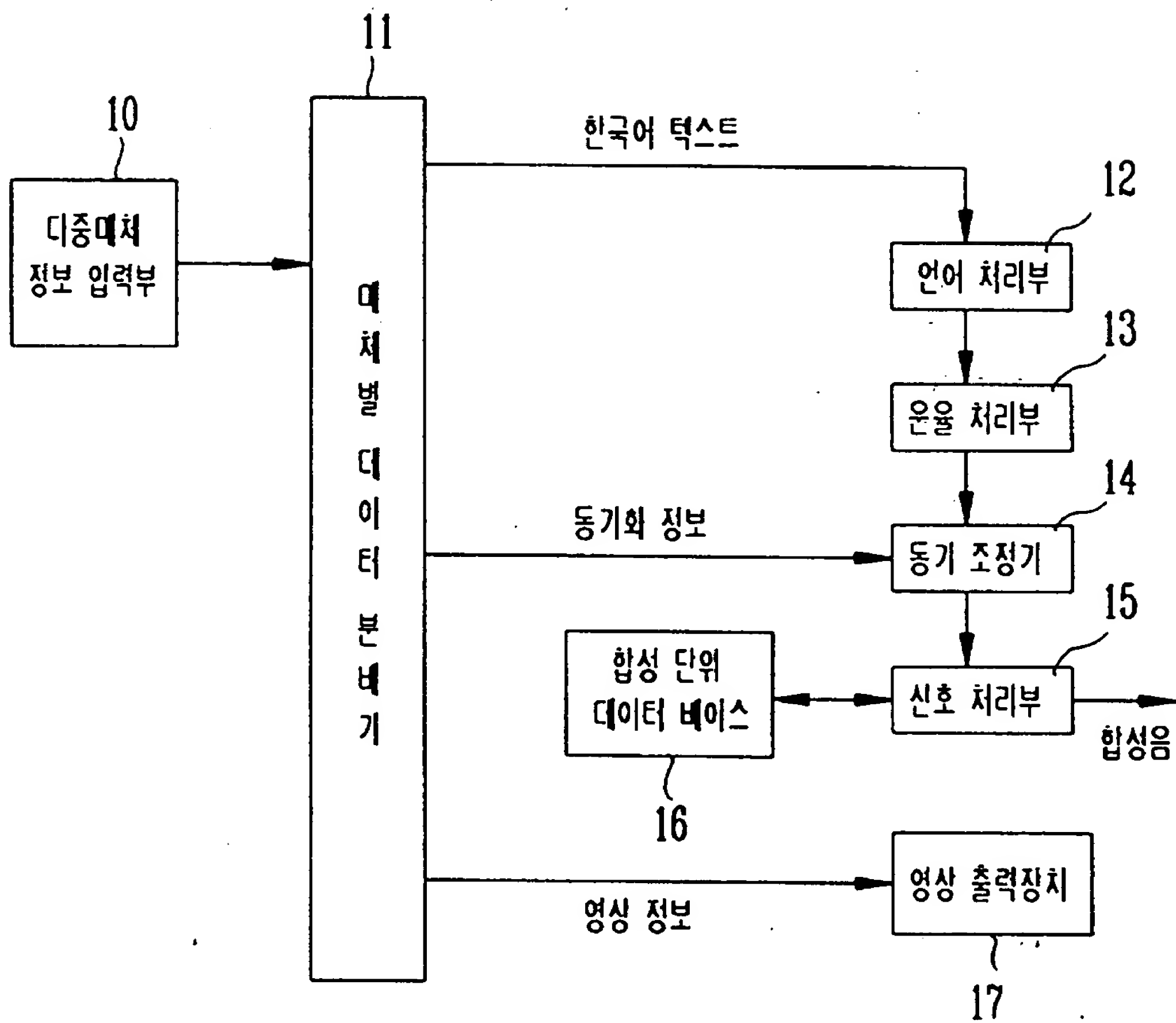
【도 1】



【도 2】



【도 3】



위 임 장

수 임 자	1 성 명	최 승 민	2 대리인 코드	441-L 112	3 전화번호	3149-7805
	1 성 명	신 영 무	2 대리인 코드	442-G 041	3 전화번호	3149-7805
	4 주 소	서울특별시 중구 순화동 1-170 삼도빌딩 4층				
사 건 의 포 시		5 출원번호	특허출원		6 출원일자	
		7 등록(항고심판) 번호			8 등록(항고심판)일자	
9 발명(고안)의 명칭		다중매체와의 연동을 위한 텍스트/음성변환기 및 그 입력 데이터 구조화 방법				
위 임 자	10 성명 또는 명칭	한국전자통신연구원 원장: 양 승택	11 주민등록번호 (외국인은 국적)	37500300	12 전화번호	(042) 860-5844
	13 주 소	대전광역시 유성구 가정동 161			우편번호	305-350
	14 사건과의 관계	출원인				
15 위 임 할 사 항	<p>(1) 상기 출원에 관한 일체행위 및 본건에 관한 출원변경, 출원분할, 포기 또는 취하, 각종의 청구, 신청, 우선권주장 또는 그 취하, 출원인 명의 변경, 기타(성명, 명칭, 인감, 주소)변경 및 개정, 거절사정 또는 보정각하 결정에 대한 항고심판청구와 그 답변 및 이에 대한 답변, 본건에 관한 특허청의 처분에 대한 소원 및 행정소송을 제기하는 권한과 본건등록의 전후에 법률 및 규칙에 따른 필요한 모든 행위를 하는 권한:</p> <p>(2) 전기사항을 처리하기 위한 복대리인의 선임 및 해임에 관한 권한.</p>					

특허법 제7조, 실용신안법 제3조, 의장법 제4조 및 상표법 제5조의 규정에 의하여 위와 같이 위임함.



1997 년 5 월 1 일

위 임 인 : 大田廣域市 儒城區 柯亭洞 161
韓國電子通信研究院
院長 梁 承

